DS-GA 1013: Project
Intelligible Principal Component Analysis

Kho, Lee
ltk224

Peeters, Jonas
jp5642

April 15, 2020

Deliverables

1. Load the data
2. Performance benchmarking
3. PCA
4. Thresholding PCA
5. Simple Principal Components
6. CPEV versus Sparsity plots
7. SCoTLASS
8. SPCA

Simple Principal Components

Vines (2000) introduced the concept of simple principal component by restricting the magnitude of the loadings to integers such as -1, 0 and 1. This approached is based on the work for Hausman (?) who introduced a integer programming solution for assembling principal components using integers from a finite set. Each element is determined sequentially as to optimize the variance explained by each component. The problem with this method is that it rarely results in orthogonality over the components.

Since it is hard to simplify principal components once they are calculated, Vines describes an alternative approach to obtaining simple components using an ortho-gonality preserving transformation on a pair of components. For each transformation of a pair of components, the variance of the data in the direction of one component is maximized while the variance in the other direction is minimized. This transformation is applied to a combination of principal components that has a high covariance. After this transformation, we exclude this pair of components for further transformations until all sensible transformations were made with the remaining components.

During the "simplicity preserving transformation" introduced by Vines (2000), we essentially orthogonally rotate and rescale two existing components d_1 and d_2. In general we can write an orthogonal rotation as $f_1 = Pd_1$ where

$$P = \begin{bmatrix} 1 & \|d_2\|_2^2\beta \\ \beta & -\|d_1\|_2^2 \end{bmatrix} \tag{1}$$

The new variance-covariance matrix V^* of the rotated data can be found as $V^* = P^T VP$ where V is the variance-covariance matrix of the data before rotation.

$$V = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} \quad \text{and} \quad V^* = P^T VP = \begin{bmatrix} 1 & \|d_2\|_2^2\beta \\ \beta & -\|d_1\|_2^2 \end{bmatrix}^T \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} \begin{bmatrix} 1 & \|d_2\|_2^2\beta \\ \beta & -\|d_1\|_2^2 \end{bmatrix}$$

The variance in the direction f_1 can be found as follows:

$$\mathrm{Var}(Xf_1) = f_1^T V f_1 = (Pd_1)^T V P d_1 = d_1^T V^* d_1 \tag{2}$$

essentially we want to find a β^* such that the variance ν in the direction of the new direction f_1 is maximized.

$$\nu = \frac{\|d_1\|_2^2 \nu_{11} + 2\beta \|d_1\| \|d_2\| \nu_{12} + \beta^2 \|d_2\|^2 \nu_{22}}{\|d_1\|^2 + \beta^2 \|d_2\|^2}$$

for which we find that ν is maximal if:

$$\beta^* = \frac{-\|d_1\| \|d_2\| (\nu_{11} - \nu_{22}) + \sqrt{\|d_1\|^2 \|d_2\|^2 (\nu_{11} - \nu_{22})^2 + 4\|d_1\|^2 \|d_2\|^2 \nu_{12}^2}}{2\|d_2\|^2 \nu_{12}}$$

References

Vines, S. K. (2000, jan). Simple principal components. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(4), 441–451. doi: 10.1111/1467-9876.00204